Statistics C206B Lecture 4 Notes

Daniel Raban

January 27, 2022

1 Spectral Methods for Mixing Times and Cutoff

1.1 Spectral methods for bounding mixing times

1.1.1 Spectral gaps and relaxation times

Today, we will take a spectral approach to mixing. If P is the transition kernel (thought of as a matrix), we want to relate the mixing time to the spectrum of P. We can show that

- 1 is an eigenvalue
- For any other eigenvalue λ, $|\lambda| \leq 1$.

Example 1.1. Can there be an eigenvalue -1? If P corresponds to the random walk on a bipartite graph, then -1 is also an eigenvalue.

Now, let's specialize to reversible chains. The detailed balance equations are

$$\pi(x)P(x,y) = \pi(y)P(y,x).$$

If π is uniform, then P is a symmetric matrix. In general, let

$$\Pi = \begin{bmatrix} \pi(1) & & & \\ & \pi(2) & & \\ & & \ddots & \\ & & & \pi(n) \end{bmatrix}.$$

Then if $A = \Pi^{1/2} P \Pi^{-1/2}$,

$$A(x,y) = \frac{\sqrt{\pi(x)}P(x,y)}{\sqrt{\pi(y)}} = A(y,x),$$

so P is similar to a symmetric matrix. This means that P admits real eigenvectors

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

We also have an orthonormal basis f_j such that

$$\frac{P^t(x,y)}{\pi(y)} = \sum_{j=1}^{n} \lambda_j^t f_j(x) f_j(y).$$

Definition 1.1. The **spectral gap** of the chain is $\gamma = 1 - \lambda_2$, where λ_2 is the second eigenvalue of P.

For complex matrices, we can look at $1 - \max_{|\lambda|<1} |\lambda|$.

Definition 1.2. The **relaxation time** is $t_{\text{rel}} := \frac{1}{\gamma}$.

Here is how we can bound t_{mix} in terms of t_{rel}:

$$t_{\text{mix}}(\varepsilon) \leq t_{\text{rel}} \log\left(\frac{1}{\varepsilon \pi_{\text{min}}}\right), \qquad \pi_{\text{min}} = \min_{x \in \Omega} \pi(x).$$

Using the second eigenfunction, we get the lower bound

$$t_{\text{mix}}(\varepsilon \geq (t_{\text{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right).$$

The second eigenvalue dictates the rate of mixing as

$$\lim_{t \to \infty} d(t)^{1/t} = \lambda_2,$$

but you usually need very large t for this to be sharp.

1.1.2 Bounds on spectral gap via contraction

Let Ω be equipped with a metric ρ, and assume that the Markov chain **contracts** with respect to ρ: For any two states $x, y \in \Omega$, if $X_0 = x$ and $Y_0 = y$, then there is a coupling of Markov chains such that

$$\mathbb{E}[\rho(X_1, Y_1)] \leq \theta \rho(x,y) \qquad \theta < 1.$$

In this case, $|\lambda| < \theta$ for all non-leading eigenvalues, so $\gamma > 1 - \theta$.

1.1.3 Using the variational characterization of eigenvalues

Consider the **Dirichlet form**, the quadratic form defined by

$$\begin{aligned}\Sigma(f) &= \frac{1}{2} \sum_{x,y} |f(x) - f(y)|^2 \pi(x) p(x,y) \\ &= f^\top (I - P) f.\end{aligned}$$

If we use the inner product

$$\langle f, g\rangle = \sum_x f(x)g(x)\pi(x),$$

then

$$\gamma = \inf_{f\perp 1} \frac{\Sigma(f)}{\|f\|_2^2}.$$

If you plug in any function here, you get an upper bound for γ and hence a lower bound for the mixing time.

1.1.4 Spectral gap in relation to bottlenecks

We can also relate the spectral gap to expansion. Recall that when we discussed bottlenecks, we had

$$\theta(S, S^c) = \sum_{x\in S, y\in S^c} \pi(x)P(x,y), \qquad \phi(S) = \frac{\theta(S,S^c)}{\pi(S)}.$$

The Markov chain version of Cheeger's inequality is

$$\frac{\phi^2}{2} < \gamma < 2\phi.$$

Example 1.2. In the simple random walk on an n-cycle, the spectral gap is $1/n^2$, so the lower bound on γ is sharp.

Example 1.3. In the random walk on the hypercube, the upper bound on γ is sharp.

1.1.5 Path coupling

Suppose the Markov chain contracts with respect to ρ and that $\rho(x,y) \geq 1$ for all $x \neq y$. Then

$$\overline{d}(t) \leq \underbrace{e^{-\alpha t}}_{\theta^t} \operatorname{diam}$$

for some $\alpha > 0$. So if the diameter is n, then the time it takes to mix is like $\log n$. In applications, θ will depend on the system size.

Example 1.4. Suppose $\theta = 1 - \frac{1}{n}$ and the diameter is n. Then

$$\overline{d}(t) \leq \left(1 - \frac{1}{n}\right)^t n,$$

so t needs to be $\approx n\log n$ to make $\overline{d}$ small.

Example 1.5. Consider the Ising model on a graph of degree at most d (density $\propto e^{-\beta \sum_{u\sim v} \mathbb{1}_{\{\sigma_u \neq \sigma_v\}}}$). We claim that if β is low enough, then the Ising model contracts. We need to come up with ρ and a coupling such that the Ising model contracts. If σ and τ are two spin configurations, we can pick $\rho(\sigma, \tau)$ to be the **Hamming distance**, the number of disagreements.

For the coupling, it suffices by the triangle inequality to consider σ and τ that only differ at one vertex v.

Step 1: Pick the same vertex to update in both configurations.

Step 2: If the vertex is not a neighbor of v, update the two chains the same (this is consistent with the marginal distributions of the Glauber dynamics).

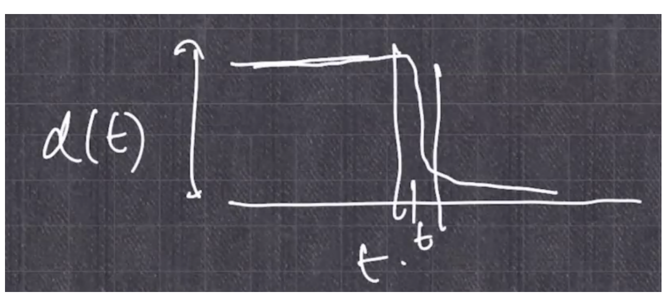
Step 3: If the vertex is a neighbor of v, the distributions of the neighbors are different between the chains. But this is like coupling two Bernoulli random variables. To couple $P_1 \sim \mathrm{Ber}(p_1)$ and $P_2 \sim \mathrm{Ber}(p_2)$, sample a single $U[0,1]$ random variable U; Then let $P_1 = \mathbb{1}_{\{U \le p_1\}}$ and $P_2 = \mathbb{1}_{\{U \le p_2\}}$.

If we update anything but a neighbor, the Hamming distance can only decrease. If we make β small enough, then we can control the probability of the Hamming distance possibly increasing when we pick a neighbor of v.

A similar argument shows a contraction for the simple exclusion process on the complete graph of size n with $n/2$ particles.

1.2 Cutoff

Cutoff is when mixing occurs abruptly.



Definition 1.3. A sequence of Markov chains indexed by n is said to exhibit **cutoff** if there exists a time t_n and a window size w_n such that $w_n = o(t_n)$ and

$$\lim_{\alpha \to -\infty} \liminf_n d(t_n + \alpha w_n) = 1,$$

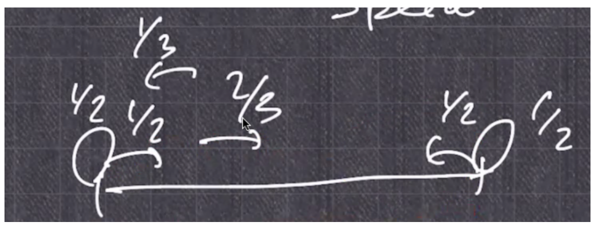
$$\lim_{\alpha \to -\infty} \limsup_n d(t_n - \alpha w_n) = 0.$$

A related notion is **precutoff**, in which

$$\sup_{0<\varepsilon<1/2} \limsup_n \frac{t_{\mathrm{mix}}(\varepsilon)}{t_{\mathrm{mix}}(1-\varepsilon)} < \infty.$$

Remark 1.1. Cutoff means that this quantity is equal to 1.

Example 1.6 (Random walk on a segment with speed)**.** Suppose we walk on the segment with drift to the right but with equal probability of staying or reflecting at the end:



The amount of time it takes to reach the right end is concentrated about $3n + O(\sqrt{n})$.

Example 1.7 (Random walk on the hypercube)**.** Last time, we used the coupon collecting problem to couple this Markov chain. Our mixing time upper bound was $n \log n$ because by that time, the two coupled chains have coalesced. In fact, this random walk exhibits cutoff at $\frac{1}{2} n \log n$. This is because at time $\sqrt{n}$, we should have collected all but one of the coupons; $\sqrt{n}$ is the order of fluctuation of the stationary measure.

Recall that

$$t_{\mathrm{mix}}(\varepsilon) \geq (t_{\mathrm{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right).$$

If we have cutoff, then the left hand side should not depend much on ε, but the right hand side blows up as $\varepsilon \to 0$. So if a Markov chain has cutoff or precutoff, the mixing time and relaxation time should not be comparable.

A necessary condition for pre-curoff (and hence cutoff) is

$$\underbrace{t_{\mathrm{rel}}}_{1/\gamma} = o(t_{\mathrm{mix}}).$$

This is the same as saying that

$$\gamma t_{\mathrm{mix}} \to \infty.$$

The converse is not true in general but is expected to be true in most "natural" examples.

Example 1.8. The random walk on the cycle does not exhibit cutoff. The mixing time is $O(n^2)$, and the spectral gap is $1/n^2$.

In general, cutoff is a high-dimensional phenomenon which occurs when you have a chain with many directions behaving independently.